

HIV Database Workshop

www.hiv.lanl.gov

seq-info@lanl.gov

Presenters: Brian Gaschen, Bette Korber, Thomas Leitner, Brian Foley, Karina Yusim

Database staff: Werner Abfalterer, Chuck Calef, Robert Funkhouser, Kristina Kommander, Dorothy Lang, John Mokili, Una Smith, Jim Szinger, Ming Zhang

Regular Contributors: John Mellors, Robert Shafer, and Christian Brander

Project Officer: James Bradac, NIAID, NIH

Carla Kuiken: www.hcv.lanl.gov

*Theoretical Biology and Biophysics, T-10
Los Alamos National Laboratory*



11th Conference
on Retroviruses and Opportunistic Infections

February 8-11, 2004 • Moscone West • San Francisco, CA

Workshop Topics

Introduction - 10 min

Sequence Database

Brian Gaschen - 30 min

Basic sequence search interface and on-the-fly alignments
Geography search interface
GeneCutter - processing nucleotide sequences
N-glycosite - infer N-linked glycosylation (if time permits)

Thomas Leitner - 30 min

HIV database alignments and subtype reference sequences
Comparing "new" sequences with our reference sequences
Build a simple neighbor joining tree through the web
Using the new RIP tool for recombination analysis

Brian Foley - 10 min

3D views of HIV macromolecular structures

Break - 15 min

Immunology Database

Bette Korber - 30 min

HIV/SIV sequence locator tool
CTL search page
Ab search page
Epitope maps

Karina Yusim - 30 min

Peptgen - list peptides for reagent development
Motifscan - find HLA anchor residues in a protein sequence
ELF - epitope location finder

Vaccine Trial Database

Brian Foley - 25 min

SHIV maps
SIV alignments
Vaccine database searches

About the instructors

- **Bette Korber** is the senior biologist on the Los Alamos HIV database project, and chief editor of the HIV immunology database.
- **Thomas Leitner** has recently replaced Dr. Carla Kuiken as chief editor for the HIV sequence database. Dr. Leitner brings to the project a decade of experience in the field of HIV evolution and sequence analysis.
- **Brian Gaschen** is the head programmer of the HIV sequence database project, and developed the HIV sequence relational database, as well as the code for many of the search and analysis tools we will be demonstrating.
- **Karina Yusim** is a postdoctoral fellow who has been involved in analysis of the data included in both the HIV sequence and immunology databases.
- **Brian Foley** joined the HIV Database team in 1995 and has extensive experience with multiple sequence alignments, phylogenetic analyses, epidemiology, the vaccine database and analyses of protein 3D structures.

Workshop Goals

- Understanding the database content, how information was obtained, and what is available
- Database searching
- Quality control tools
- Tools for analyses

The screenshot shows the HIV Sequence Database homepage in a web browser. The browser's address bar displays <http://www.hiv.lanl.gov/content/hiv-db/mainpage.html>. The page features a sidebar on the left with navigation links under categories like 'Databases', 'Publications', 'Sequence DB', and 'HCV Databases'. The main content area includes a 'News' section with several announcements, an 'About this website' section, 'Programs and tools', 'Alignments', 'Compendia', and 'Links' sections. The Los Alamos National Laboratory logo is visible in the bottom right corner of the browser window.

HIV Sequence Database

Search Site

Databases

- Sequence DB
- Resistance DB
- Immunology DB
- Vaccine trials DB

Publications

- FAQ
- Alignments
- Tutorials
- Reviews
- Compendia
- Links

Sequence DB

- Search DB
- Tools
- HIV-BLAST
- Recombination
- Syn-Nonsyn
- Hypermut
- PCOORD
- SUDI
- TreeMaker
- Geography
- N-Glycosite
- 3D Structure
- GeneCutter
- Rip 1.9 Beta

HCV Databases

[Disclaimer/Privacy](#)

News

- The 11th Conference on Retroviruses and Opportunistic Infections will be held February 8-11, 2004 at Moscone West, San Francisco, CA. The database crew will provide two workshops on how to use the HIV and SIV databases. Space is limited, and prior registration is necessary through the conference. 29 January 2004
- The web site is now searchable. 11 December 2003
- The 11th International Workshop on HIV Dynamics and Evolution will be held April 29 - May 2, 2004 at the S  staholm Conference Center outside of Stockholm, Sweden. 30 November 2003
- The 2002 Sequence Alignments are now available. 30 November 2003
- Improved versions of the Primalign and Sequence Locator tools are available. 20 November 2003
- The Hepatitis C sequence database is now publicly accessible. It is based on the HIV database, and offers a similar set of tools, analysis options, and annotations. The two databases will post permanent links to each other's websites (see button above), and will share resources, tools and programs. An all-new HCV immunology database will become available in the winter or spring of 2004. 05 November 2003

About this website:

- [Overview of the site](#)
- [Frequently asked questions](#)

Programs and tools:

- [Search interface](#): retrieve sequences based on all HIV database search fields. HIV-1 sequences can be aligned and clipped. This interface combines the HIVMAP and DBSearch interfaces. On the search interface webpage, a link to the old search interface is still available.
- [Tools for working with sequences](#): GeneCutter, SeqConvert, Gapstrip, Motifscan, Primalign, Epilign, HXB2 Numbering, SeqPublish, HIV-BLAST, sequence format conversion, N-GlycoSite
- [Programs for sequence analysis](#): RIP (Recombination), SNAP (Syn-Nonsyn), VESPA (Signature patterns), Hypermut, PCOORD, and more...
- [Other programs](#) for analysis of HMA data and optical density data
- [Links to external programs](#): phylogenetics, recombination, subtyping, multiple alignment, sequence submission
- [Tutorials](#) on subtype and CRF nomenclature, sequence quality control and tree building

Alignments:

- [Complete alignments](#) of all genes (nt and aa) and complete genomes (nt only)
- [Subtype reference alignments](#) for use in trees, subtype comparisons and recombination research
- [Consensus and Ancestral sequences](#) for M_GROUP and subtypes

Compendia:

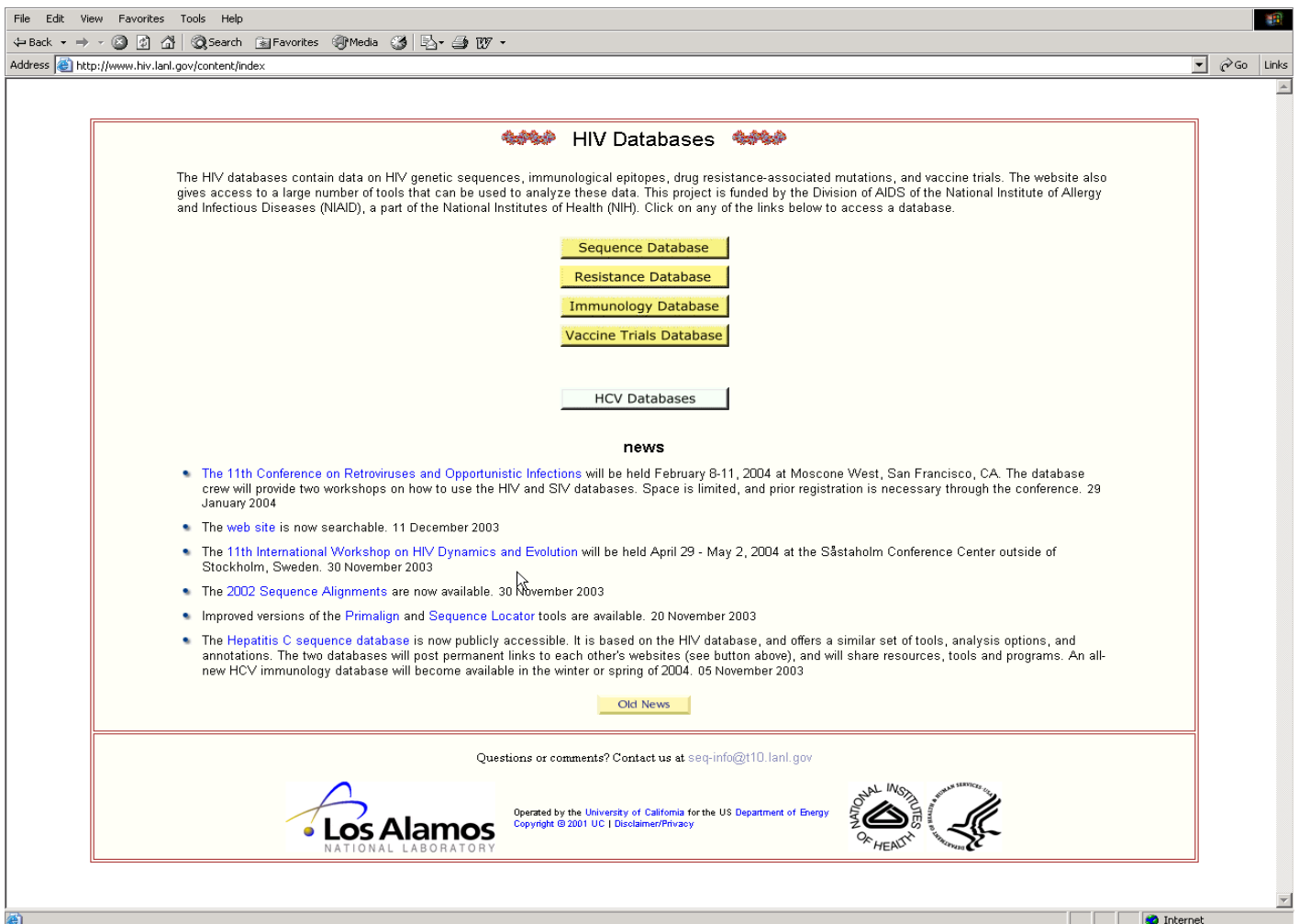
- [Print \(PDF\) or order a copy](#) of our compendia "Human retroviruses and AIDS"
- [Reviews published in the Sequence and Immunology Compendia](#)
- [How to refer to the compendia](#) in a publication

Links:

- [Los Alamos Immunology Website](#), our sister site, houses a huge searchable collection of HIV immunological epitopes
- [Los Alamos Drug Resistance Database](#) contains information about anti-HIV drugs and drug-resistance-conferring mutations
- [Los Alamos HIV/SIV Vaccine Trial Database](#) A database of vaccine trials, including design, type of vaccine used, results, etc.
- [Other HIV/AIDS sites](#) for more software and information on HIV/AIDS

The HIV databases

- HIV Sequence database – founded 1986, G. Myers
 - Relational database, data from GenBank with added fields from the literature
 - Alignments – align indels and reduce multiple sequences per person
 - Annual hard copy and reviews
 - Web search interfaces: subtype, phenotype, geographic, sampling year...
 - Analysis tools
- HIV Immunology database – founded 1995, B. Korber
 - Comprehensive HIV epitope database , 300-400 papers a year
 - Integrate HIV immunological and sequence data
 - Annual hard copy and reviews
 - Web search interfaces: epitope, protein, HLA type, immunogen, keywords
 - Analysis tools for immunologists
- HIV Drug Resistance database, founded 1997, J. Mellors
 - A searchable web listing of drug resistance mutations and literature links, updated annually by Dr. Mellors
- HIV Vaccine database, founded 2003, J. Mokili
 - A searchable relational database of published primate vaccine trials



Search Interface

■ Help

- Tips at the top of the page are often overlooked
 - Ranges, operators, wildcards, logical groupings
- Field names are clickable, also mouse-overs
 - Example: “Sampling country” gives two-letter ISO country codes

■ Searches

- Searches are case-insensitive
- Records are searchable through sequence, patient, genomic region, or publication information
- First seven fields will appear in search results page by default
- A “*” in a textbox will cause that field to be included in the results page
- Patient information (Infection year, Infection country) is different than sequence information (Sampling year and Sampling country)

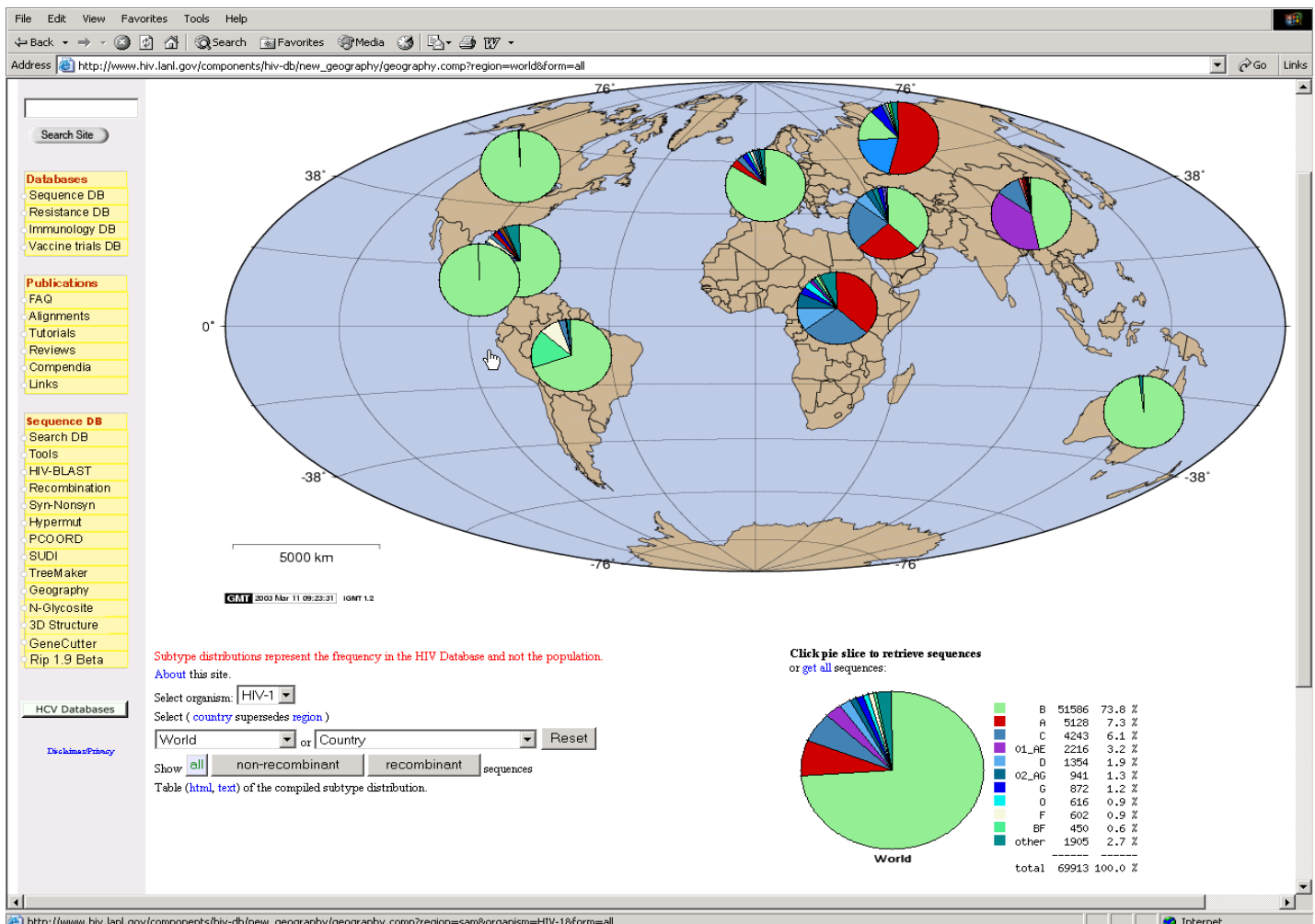
■ Results

- Can select not aligned, or aligned based on multiple pair wise alignments – alignments are good, but still need hand editing for an optimal alignment
- Select all or a subset of sequences for download
- Sequences can be re-ordered by clicking on fields at the top of the page

The screenshot shows the HIV Sequence Database search interface in a web browser. The browser's address bar displays the URL: http://www.hiv.lanl.gov/content/hiv-db/combined_search/search. The interface includes a sidebar with navigation links for Databases (Sequence DB, Resistance DB, Immunology DB, Vaccine trials DB), Publications (FAQ, Alignments, Tutorials, Reviews, Compendia, Links), and Sequence DB (Search DB, Tools, HIV-BLAST, Recombination, Syn-Nonsyn, Hypermut, PCOORD, SUDI, TreeMaker, Geography, N-Glycosite, 3D Structure, GeneCutter, Rip 1.9 Beta). The main search area contains various input fields: Accession number, Subtype (dropdown), Sequence name, include recombinants (checkbox), Sampling country, Organism (dropdown), Sampling year, Sequence length, Patient code, Infection country, Risk factor (dropdown), Infection year, Coreceptor (dropdown), Phenotype (dropdown), Author(s), Pubmed/Medline ID, Geographic region, and Other fields (dropdown). A red note states: "Use this option to search based on the HIV-DB internal alignment (presently this option restricts the search to HIV-1):". Below this, there is a Genomic region dropdown and a checkbox for "Include fragments of minimum length 100". At the bottom, there is a "List 100 records per page" option, "Search" and "Reset" buttons, and a link to "Old Search Interfaces". The footer includes the Los Alamos National Laboratory logo, contact information (seq-info@t10.lanl.gov), and logos for the University of California and the National Institutes of Health.

Geography Tool

- Another way to search/download sequences is by geographic region or country
- Results are biased as they show only the sampled individuals, not the true subtype distribution for a region's population
- Results are selectable as in the search interface



Gene Cutter

- Useful for sequencing labs, particularly for rapid processing of new sets of full length genomes
- Cut out genes and proteins from aligned sets of DNA sequences
- Sequences do not need to be codon aligned – results can be codon-aligned on the fly with generally good results
- Currently, sequence alignments must contain HXB2 as a reference for the program to function

The screenshot shows the Gene Cutter web interface within a Netscape browser window. The address bar shows the URL: http://www.hiv.lanl.gov/content/hiv-db/GENE_CUTTER/cutter.html. The page has a sidebar on the left with navigation links for 'Databases' (Sequence DB, Resistance DB, Immunology DB, Vaccine trials DB), 'Publications' (FAQ, Alignments, Tutorials, Reviews, Compendia, Links), and 'Sequence DB' (Search DB, Tools, HIV-BLAST, Recombination, Syn-Nonsyn, Hypermut, PCORD, SUDI, TreeMaker, Geography, N-Glycosite, 3D Structure, GeneCutter, Rip 1.9 Beta). The main content area is titled 'GENE CUTTER' and contains the following text:

Gene Cutter is a tool that clips pre-defined coding regions from a nucleotide alignment, then codon aligns and provides translations of the cut regions. To define boundaries of genes or domains of interest, and to codon align the sequences, Gene Cutter uses the coordinates from a reference sequence (HXB2). Thus GeneCutter requires the proper HXB2 sequence ([K03455](#)) to be in the input nucleotide alignment. It uses HXB2 to set the proper reading frame for the rest of the alignment by slightly rearranging the sequence alignment so that HXB2 is in frame.

So, in order for Gene Cutter to do its job, your alignment must contain HXB2.

Remember: Your output alignment will only be as good as your input alignment.

Note: In some HIV sequences an insertion will be compensated for within a short distance by a deletion, or vice versa. As these frame shifts may not inactivate the protein, if a compensating mutation is within 5 amino acids of an initial frameshift, the frame-shifted reading frame is left intact. Otherwise, the frame shift is marked with the hash symbol (#), and the translation is continued in the correct, typical reading frame beyond the offending codon. Stop codons are marked by a dollar sign (\$).

Input options:
Select the region of the alignment you would like to extract. Note: The HXB2 sequence contained in your alignment **does not** need to cover all of the selected region to be cut correctly.

Enter your **alignment containing HXB2** here.
Note: Your sequence does not need to be named HXB2.

Input alignment format:

Output options:
Output alignment format:

☒ Codon align the region
☐ View region in browser (do not download)
☐ Translate region to amino acids (translates all codons containing [IUPAC/IUB multistate characters](#) to "X"; codons containing "-" are translated to either "." or "#")
☐ Translate region to amino acids (translates codons containing [IUPAC/IUB multistate characters](#) that are involved in silent substitutions, "X" otherwise; codons containing "-" are translated to either "." or "#")

N-GLYCOSITE

- Tracks of patterns of N-linked glycosylation site (N-X-[ST]) change in sequences
- INPUT: A sequence alignment of interest
- OUTPUT:
 - Tallies of numbers of Ngly sites in each sequence
 - Highlighted Ngly sites
 - Graphics illustrating frequency of Ngly patterns in the alignment, and in sub-regions of the alignment
 - Frequencies of different patterns of X and Y in N-X-[ST]-Y



Sequence alignments

- Originally based on iterations of manual and HMM alignments
- Yearly updates using HMM and manual corrections
- Full length genomes updated throughout the year
- Alignments are in reading frame (codon aligned)
- Alignments non-redundant
- Compendia alignments show fewer sequences than web version
- Reference alignments contain up to four representatives
- Protein alignments may contain frameshift compensations
- Subtype consensus with ties resolved, as well as maximum likelihood ancestors, are available for reagent production

The screenshot shows the HIV Sequence Database website. The main heading is "2002 HIV and SIV alignments". Below it, there is a dropdown menu for "Alignment format" set to "FASTA". A link points to "Click here for information on color-coded protein alignments." There are "Get Alignment" and "reset" buttons. A note explains that protein alignments are based on both nucleotide and translated amino acid sequences. A table lists various regions and their corresponding alignment options (DNA or Protein).

Region	HIV-1/SIVcpz	HIV-2/SIVsmm	Other SIV
Genome	<input type="radio"/> DNA	<input type="radio"/> DNA	<input type="radio"/> DNA (includes HIV-1 and HIV-2 sequences)
LTR	<input type="radio"/> DNA	<input type="radio"/> DNA	<input type="radio"/> DNA
GAG	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
POL	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
ENV	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
VIF	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
TAT	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
REV	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
VPU/VPX	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
VPR	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
NEF	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein

Questions or comments? Contact us at seq-info@t10.lanl.gov

Los Alamos NATIONAL LABORATORY

Operated by the University of California for the US Department of Energy
Copyright © 2001 UC | Disclaimer/Privacy

NATIONAL INSTITUTE OF HEALTH

INTERNET

Comparing "new" and database sequences

- As an example, eight "new" *env* sequences have been aligned to subtype references (A-K) and sequences from previous search (Japanese *env*) using HMM, Genecutter & Se-AI
- TreeMaker produces a Neighbor Joining tree for a "quick-and-dirty" comparison
- TreeMaker is based on DNADIST & NEIGHBOR in the PHYLIP package
- HIV-BLAST is an option for looking for highly similar sequences or possible contamination

NEIGHBOR TREEMAKER

Neighbor TreeMaker takes a sequence alignment, converts it to [PHYLIP](#) format, runs it through the PHYLIP programs **Dnadist** (Distance Matrix program), and **Neighbor** (treefile generator), then displays a tree.

The **Dnadist** program reads in nucleotide sequences and writes an output file containing the distance matrix. PHYLIP gives options concerning the model used to calculate distances. This interface uses ML, the model used in PHYLIP's maximum likelihood phylogeny program DNAML. This model incorporates different rates of transition and transversion, and also allows for different frequencies of the four nucleotides. [PHYLIP copyright/reference info.](#)

Disclaimer: This interface only offers very basic, 'quick-and-dirty' phylogenetic analysis. More in-depth analysis is usually needed. For more information see the [TreeMaker Tutorial](#).

1. Submit Alignment.

Please paste your alignment into the submission box below and **indicate the format**. This interface only accepts nucleotide sequences.

Paste your alignment here:

Current Format: [Sample Input](#)

2. Set Program Parameters.

You can tune **Neighbor TreeMaker** by adjusting two parameters.

a. specify the number of the **outgroup** sequence in your alignment (to be the root of your tree; default is the first sequence):

b. specify the **transition/transversion ratio**

3. Choose Program Output.

The tree can be presented as a

☒ Phenogram or as a

☐ Radial Tree

(Once the tree has been generated, you can view the text treefile and outfile.)

4. Execute Program.

Please double-check the submission form if you get any abnormal output, and [contact us](#) if the problem persists.

Recombination Analysis

- Many methods and programs exist to investigate potential recombination
 - <http://bioinf.man.ac.uk/~robertson/recombination/>
- Investigating recombination requires many steps
- A new version of RIP is available at HIV db
 - Automatic alignment
 - Selection of background sequences
 - Different window/gap handling options
 - Graphic & table output

The screenshot shows a web browser window displaying the HIV Sequence Database RIP 2.0 Beta Version interface. The page has a purple header with the title 'RIP 2.0 Beta Version'. Below the header, there is a section for 'QUERY SEQUENCE' with a dropdown menu for 'Format of your query sequence' set to 'FASTA', a text input field for 'Upload your query sequence file ...', and a 'Browse...' button. There is also a 'Sample Input' button and a text area for 'or paste your query sequence here:'. Below this is a section for 'BACKGROUND ALIGNMENT' with radio buttons for 'Use subtype consensus sequences as background', 'Include 01_AE consensus', 'Exclude 01_AE consensus', 'Select custom background from subtype consensus and representative sequences', and 'Use your own alignment: query (first sequence) aligned to background'. There are input fields for 'Upload your alignment ...' and 'or paste your background here ...', each with a 'Browse...' button. Below this is a section for 'RIP SETTINGS' with dropdown menus for 'Window size' (set to 200) and 'Significance threshold' (set to 90%), and radio buttons for 'Output format' (set to Graphical). There are 'Reset' and 'Run' buttons at the bottom of the settings section. The left sidebar contains a 'HIV Sequence Database' logo, a 'Search Site' button, and a list of 'Databases' (Sequence DB, Resistance DB, Immunology DB, Vaccine trials DB), 'Publications' (FAQ, Alignments, Tutorials, Reviews, Compendia, Links), and 'Sequence DB' (Search DB, Tools, HIV-BLAST, Recombination, Syn-Nonsyn, Hypermut, PCO ORD, SUDI, TreeMaker, Geography, N-Glycosite, 3D Structure, GeneCutter, Rip 1.9 Beta). The footer includes the Los Alamos National Laboratory logo, contact information, and logos for the National Institutes of Health and the University of California.

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print

Address http://hiv-dev.lanl.gov/content/hiv-db/RIPPER2/rip_test.html Go Links

HIV Sequence Database

Search Site

Databases

- Sequence DB
- Resistance DB
- Immunology DB
- Vaccine trials DB

Publications

- FAQ
- Alignments
- Tutorials
- Reviews
- Compendia
- Links

Sequence DB

- Search DB
- Tools
- HIV-BLAST
- Recombination
- Syn-Nonsyn
- Hypermut
- PCO ORD
- SUDI
- TreeMaker
- Geography
- N-Glycosite
- 3D Structure
- GeneCutter
- Rip 1.9 Beta

HCV Databases

- Update Database
- Disclaimer/Privacy

RIP 2.0 Beta Version

This is the sequence submission page for the new beta version of RIP. In this version of RIP your query sequence is automatically aligned to the background alignment. The alignment so produced may be downloaded from the results page. The site is under development and you will notice some of the options have been deactivated. The output of this program should be interpreted with caution. Please [send us](#) bug reports or suggestions.

QUERY SEQUENCE

Format of your query sequence: FASTA

Upload your query sequence file ... Browse...

Sample Input

or paste your query sequence here:

BACKGROUND ALIGNMENT

☒ Use subtype consensus sequences as background

☐ Include 01_AE consensus ☐ Exclude 01_AE consensus

☐ Select custom background from subtype consensus and representative sequences

☐ Use your own alignment: query (first sequence) aligned to background

Upload your alignment ... Browse...

or paste your background here ...

RIP SETTINGS

Window size: 200 Significance threshold: 90%

Output format: ☒ Graphical ☐ Tabular ☐ Both

Reset Run

Questions or comments? Contact us at seq-info@t10.lanl.gov

Los Alamos NATIONAL LABORATORY

Operated by the University of California for the US Department of Energy
Copyright © 2001 UC | [Disclaimer/Privacy](#)

NATIONAL INSTITUTES OF HEALTH

Internet

HIV and SIV Protein

3D Structures


- Structures Determined by X-Ray Crystallographic and NMR Methods
 - HIV and SIV Structural proteins (RT, Protease, Env Core)
 - RNA Secondary Structures (TAR, PSI, RRE)
- Models of non-crystalized proteins
 - Env Core with Variable Loops added
- Tutorials and Reviews
 - Env, RT, Protease
- Links to Free 3D Viewer Software
 - CHIME/Protein Explorer
 - RasMol
 - Visual Molecular Dynamics

HIV 3D STRUCTURES INDEX - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print Mail Print Mail

Address http://www.hiv.lanl.gov/content/hiv-db/STRUCTURE/INDEX.HTML Go Links



Search site

Databases

- Sequence DB
- Resistance DB
- Immunology DB
- Vaccine trials DB

Publications

- FAQ
- Alignments
- Tutorials
- Reviews
- Compendia
- Links

Sequence DB

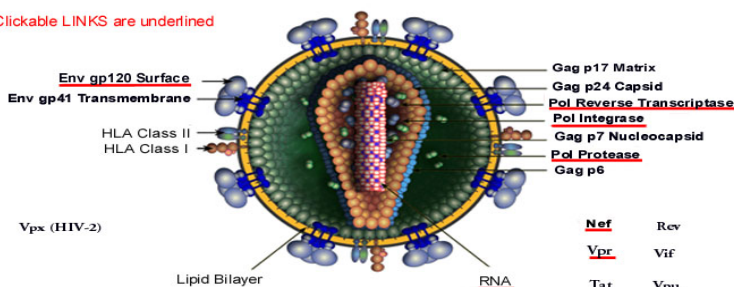
- Search DB
- Tools
- HIV-BLAST
- Recombination
- Syn-Nonsyn
- Hypermut
- PCOORD
- SUDI
- TreeMaker
- Geography
- N-Glycosite
- 3D Structure
- GeneCutter
- Rip 1.9 Beta

HCV Databases

Disclaimer/Privacy

Index to 3D Views of HIV Macromolecular Structures

Clickable LINKS are underlined



Env gp120 Surface
Env gp41 Transmembrane
HLA Class II
HLA Class I
Vpx (HIV-2)
Lipid Bilayer
RNA

Gag p17 Matrix
Gag p24 Capsid
Pol Reverse Transcriptase
Pol Integrase
Gag p7 Nucleocapsid
Pol Protease
Gag p6

<u>Nef</u>	Rev
<u>Vpr</u>	Vif
Tat	Vpu


Index of 3D Structure Tools, Tutorials and Links

Mature HIV-1 Virion
Image courtesy of Lou Henderson
henderson@neicf.gov


A review of the Structural Biology of HIV
J Mol Biol 1999 Jan 8;285(1):1-32
Structural biology of HIV.
Turner BG, Summers MF.

In order to view and interact with the tutorials on 3D Structures of HIV proteins contained within these WWW pages, it is recommended that you use Netscape Communicator versions 4.5 - 4.76 (many versions of Internet Explorer also work, but some CHIME pages do not work with IE.) and **required** that you use the free MDL Chemscape Chime www browser plug-in.
You can Test your CHIME installation here.

Questions or comments? Contact us at seq-info@t10.lanl.gov



Operated by the University of California for the US Department of Energy
Copyright © 2001 UC | Disclaimer/Privacy



HIV/SIV Sequence Locator Tool

- Rapidly returns position numbers of an HIV or SIV DNA or protein sequence fragment relative to the HXB2r or SMM239 reference strains.
- Such numbers are often included in the literature, and are often incorrect
- Marks the location of the sequence on an HIV map
- For DNA sequences, a translation is provided
- Can be used for input into the search interface, to align a new sequence you have generated with the database set.

The screenshot shows the HIV/SIV Sequence Locator web application. The browser address bar shows the URL: http://www.hiv.lanl.gov/content/hiv-db/LOCATE_SEQ/locate.html. The page title is "HIV/SIV SEQUENCE LOCATOR".

Left Sidebar:

- HIV Sequence Database**
- Databases:** Sequence DB, Resistance DB, Immunology DB, Vaccine trials DB
- Publications:** FAQ, Alignments, Tutorials, Reviews, Compendia, Links
- Sequence DB:** Search DB, Tools, HIV-BLAST, Recombination, Syn-Nonsyn, Hypermut, PCOORD, SUDI, TreeMaker, Geography, N-Glycosite, 3D Structure, GeneCutter, Rip 1.9 Beta
- HCV Databases**
- [Disclaimer/Privacy](#)

Main Content Area:

Paste your sequence in the box below, or use the browse button to select a file that contains sequence(s) to upload. You can specify that your input sequence(s) are HIV or SIV by selecting the appropriate choice from the list on the right, or you can let the program decide (default). [Details](#).

Let program decide
HIV
SIV

Single sequence
Multiple sequences

If you submit multiple sequences then indicate that here -->

☐ Examine reverse complement of sequence

Sample Input

Browse...

Reset RUN SEQ LOCATOR

Input Options: Default is "Single sequence". See options in table below. You can mix nucleotide and amino acid sequences in your input.

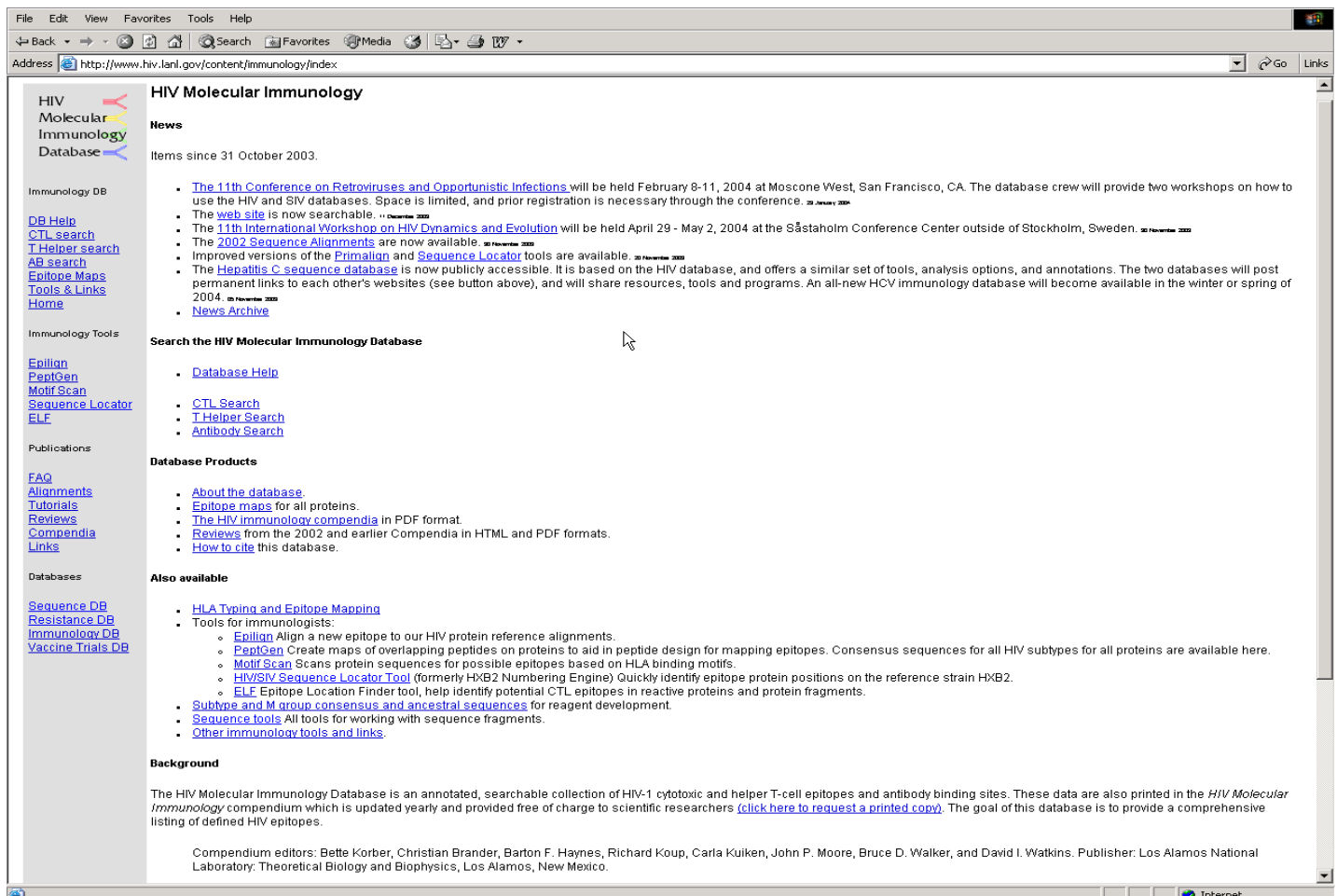
Single sequence	Multiple sequences
<p>1. Free format. Sequence can contain carriage returns, spaces, dashes, and other characters. The program will remove all non-letter characters like returns and spaces before processing.</p> <p>Example input:</p> <pre>act gatgc---tcacgtatcg xactt agn tagtcga</pre> <p>will be treated as</p> <pre>actgatgctcactgatacgacttagtagtcga</pre> <p>2. Fasta format. Example:</p> <pre>>MySeq tgtgatcgtagtgctatcatgctgtgacgagcgatcg</pre>	<p>To submit multiple sequence the user must select the "Multiple Sequences" option. The program recognizes two multiple sequence input options:</p> <p>1. Fasta format. Example:</p> <pre>>seq1 LAEEVVVIRSENFDTNAKTIIVQLN ESVREINCTRPNNN >seq2 GPGRAFYTTCGIIIGDIRQAHG >seq3 VTKLREQFKN-KTIVFNQSSGGD</pre> <p>The program recognizes this format by the ">" character. All non-letter characters will be removed, so it is OK if your sequence contains returns (e.g. seq1) or gaps (seq3).</p> <p>2. Raw sequence, one sequence per line with a carriage return between sequences. Example:</p> <pre>LAEEVVVIRSENFDTNAKTIIVQLN ESVREINCTRPNNN GPGRAFYTTCGIIIGDIRQAHG VTKLREQFKN-KTIVFNQSSGGD</pre>

HIV or SIV? You can specify that your input sequence(s) are HIV or SIV by selecting the appropriate choice in the list, or you can let the program decide.

Reverse complement? If you check this box the sequence locator creates a sequence that is the reverse complement of your input sequence and tries to locate it in the HXB2 or SMM239 genomes. If the match score is better for the reverse complement than for the original sequence you will be informed.

Immunology Database Overview

- HIV T-Cell (CTL, T-helper) and Antibodies (Ab)
- Types of data recorded
 - Epitope sequence and location: HXB2 numbering, subtype
 - Immunogen
 - Host HLA or MHC, and Ab isotype
 - Notes summarize main findings
- Contents: data from 1985 through 2002
 - 2300 CTL entries
 - 600 T-helper entries
 - 1100 Ab entries



Immunology Database: Search

■ T Cells

- ☐ Cytotoxic T Lymphocytes (CTL)
- ☐ Helper T Lymphocytes (T-helper)
- ☐ Biological distinction between CTL and T-helper is not always obvious
- ☐ Organization is identical for CTL and T-helper
- ☐ One reference per entry

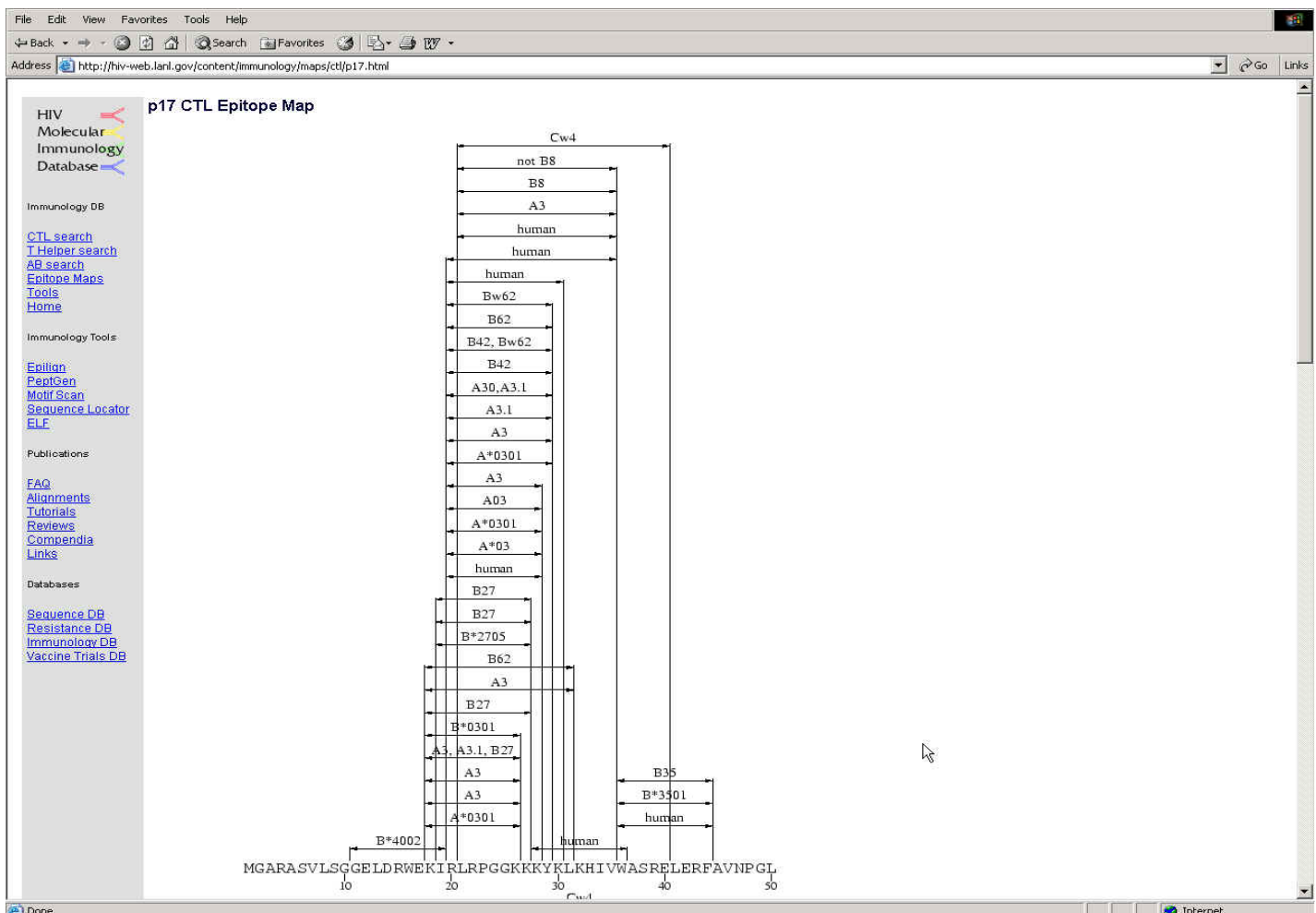
■ B Cells (Antibodies)

- ☐ One entry for each monoclonal antibodies
- ☐ Many references per entry (up to 100)

The screenshot shows a web browser window with the address http://www.hiv.lanl.gov/content/immunology/ctl_search. The page title is "HIV Immunology CTL Search". On the left is a navigation menu with links for HIV, Molecular Immunology Database, Immunology DB, DB Help, CTL search, T Helper search, AB search, Epitope Maps, Tools & Links, Home, Immunology Tools, Epitope, PeptGen, Motif Scan, Sequence Locator, ELE, Publications, FAQ, Alignments, Tutorials, Reviews, Compendia, Links, Databases, Sequence DB, Resistance DB, Immunology DB, and Vaccine Trials DB. The main search area contains several dropdown menus: HIV Protein (with sub-sections for Proteins with defined epitopes and Reactive proteins with undefined epitopes), Epitope, Immunogen, Vaccine details (with a note "If Immunogen is Vaccine, additional search details"), Species, HLA, Author, and Keywords. Each dropdown menu has a list of options, including "-ALL-", "-NULL-", and various specific identifiers. At the bottom of the search area are "Search" and "Reset" buttons. The footer of the page includes the email immuno@t10.lanl.gov and the Los Alamos National Laboratory logo.

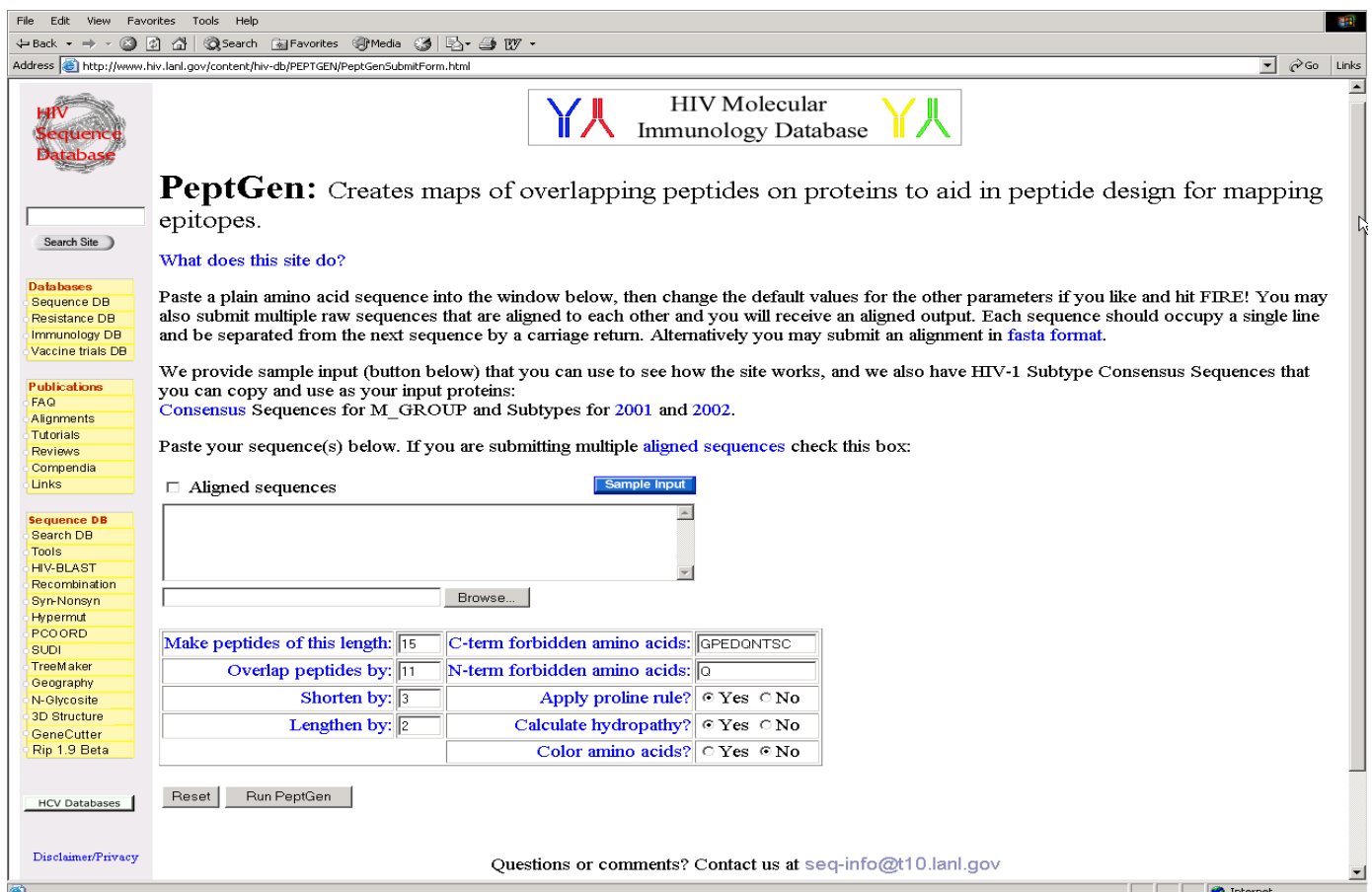
Immunology Database: Additional Information

- All entries for a reference
- Medline links to papers
- Epitope Tables
- Epitope Maps
 - Unique species/HLA for T cell epitopes
 - MAb name, species code for Ab
- Epitope Alignments
 - Extracted from HIV-sequence database, includes subtype, country and year of sampling



PeptGen

- Creates maps of overlapping peptides on proteins to aid in peptide design for mapping epitopes
- Consensus sequences for all HIV subtypes for all proteins are available
- Use alignments to design comparable sets of peptides (for example, to compare clades)
- INPUT
 - ☐ Query sequence or aligned sequences
 - ☐ Desired length of peptides, peptide overlap, forbidden C- and N-terminal amino acids
- OUTPUT
 - ☐ Maps of overlapping peptides (forbidden amino acids are taken into account)
 - ☐ Highlighted forbidden amino acids
 - ☐ Hydropathicity scores for the peptides are available



The screenshot shows the PeptGen web interface in a browser window. The address bar shows the URL: <http://www.hiv.lanl.gov/content/hiv-db/PEPTGEN/PeptGenSubmitForm.html>. The page features a sidebar with navigation links for HIV Sequence Database, Publications, and Sequence DB. The main content area includes a description of PeptGen, instructions on how to use the tool, and a form for submitting sequences and parameters. The form includes fields for sequence input, peptide length, overlap, and various options like 'C-term forbidden amino acids' and 'Calculate hydropathy?'. A 'Run PeptGen' button is at the bottom of the form.

PeptGen: Creates maps of overlapping peptides on proteins to aid in peptide design for mapping epitopes.

What does this site do?

Paste a plain amino acid sequence into the window below, then change the default values for the other parameters if you like and hit FIRE! You may also submit multiple raw sequences that are aligned to each other and you will receive an aligned output. Each sequence should occupy a single line and be separated from the next sequence by a carriage return. Alternatively you may submit an alignment in [fasta format](#).

We provide sample input (button below) that you can use to see how the site works, and we also have HIV-1 Subtype Consensus Sequences that you can copy and use as your input proteins:
[Consensus Sequences for M_GROUP](#) and Subtypes for [2001](#) and [2002](#).

Paste your sequence(s) below. If you are submitting multiple [aligned sequences](#) check this box:

☐ Aligned sequences [Sample Input](#)

Make peptides of this length:	<input type="text" value="15"/>	C-term forbidden amino acids:	<input type="text" value="GPEDQNTSC"/>
Overlap peptides by:	<input type="text" value="11"/>	N-term forbidden amino acids:	<input type="text" value="Q"/>
Shorten by:	<input type="text" value="3"/>	Apply proline rule?	<input checked="" type="radio"/> Yes <input type="radio"/> No
Lengthen by:	<input type="text" value="2"/>	Calculate hydropathy?	<input checked="" type="radio"/> Yes <input type="radio"/> No
		Color amino acids?	<input type="radio"/> Yes <input checked="" type="radio"/> No

Questions or comments? Contact us at seq-info@t10.lanl.gov

HLA Binding Motif Scanner : **MotifScan**

- Finds HLA anchor motifs within protein sequences for specified HLA genotypes, serotypes, or supertypes
- HLA anchor motif dictionaries are available on line
- Main motif and supermotif sources:
 - SYPHEITHI Database, Rammensee *et al.* www.syfpeithi.de
 - HLA Facts Book, Marsh *et al.* 2000
 - Sette & Sidney, *Immunogenetics* **50**:201-212, 1999
- INPUT:
 - User defined query sequence or aligned sequences, or reference set
 - Selected HLA anchor motifs, or user defined motif
 - The user defined motif function could be used to search for other patterns of interest in sequences
- OUTPUT:
 - Anchor residue positions are highlighted in the query sequence
 - Potential epitopes and positions are listed
 - Output can be downloaded as text, convenient for further analysis

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media Print W

Address http://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan Go Links

HIV
Molecular
Immunology
Database

Immunology DB

DB Help
CTL search
T Helper search
AB search
Epitope Maps
Tools & Links
Home

Immunology Tools

Epilign
PeptGen
Motif Scan
Sequence Locator
ELE

Publications

FAQ
Alignments
Tutorials
Reviews
Compendia
Links

Databases

Sequence DB
Resistance DB
Immunology DB
Vaccine Trials DB

HLA Binding Motif Scanner

Use this page to find HLA anchor residue motifs within protein sequences for specified HLA genotype, serotypes or supertypes. Refer to the [Help](#) page for more information.

Please enter your search criteria

Genotype	Serotype	Supertype
A*01	A1	A1
A*0101	A2	A2
A*0201	A3	A3
A*0202	A11	A24
A*0204	A24(9)	B7
A*0205	A26(10)	B27
A*0206	A29(19)	B44

[HLAs](#)

[Motif Source](#) ☒ Marsh2000 ☒ SYFPEITHI ☒ Others

[Motif Length](#) ☐ 8 ☒ 9 ☐ 10 ☐ 11

[Custom Motif](#)

Search **Reset**

Data dictionaries

[View](#) or [download](#) the HLA genotype/serotype dictionary.

[View](#) or [download](#) the HLA genotype/motif dictionary.

[View](#) or [download](#) the HLA supertype dictionary.

immuno@t10.lanl.gov

Operated by the [University of California](#) for the US [Department of Energy](#)
Copyright © 2003 UC | [Disclaimer/Privacy](#)

ELF: E pitope L ocation F inder

- Helps identify potential CTL epitopes in reactive peptides

- INPUT:

- ☐ Reactive peptide sequence
- ☐ Full protein sequence that peptides were based on
- ☐ Patient's HLA information – genotype or serotype

- OUTPUT:

- ☐ If HLA serotype is submitted, associated HLA genotypes are given
- ☐ Potential epitopes in reactive peptides, based on anchor motifs
- ☐ Maps of all HIV epitopes for all HIV proteins, highlighting epitopes that use the patient's HLA presenting molecules
- ☐ Location of the query peptide according to the HXB2 reference strain
- ☐ Alignment of query peptide against reference database alignments
- ☐ All known CTL HIV epitopes contained in the query peptide
- ☐ Epitopes presented by the patient's HLA presenting molecules, that are potentially experimentally missed because of amino acid differences between the previously defined epitopes and the query strain

File Edit View Favorites Tools Help

Address http://www.hiv.lanl.gov/content/hiv-db/ALABAMA/epitope_analyzer.html

HIV Molecular Immunology Database

This is a beta-test version of the HIV Molecular Immunology Database "ELF" site. Please email comments, suggestions, and problems to ccc@t10.lanl.gov or btg@t10.lanl.gov.

What does this site do?

Instructions: Fill in as many of the boxes in the table below as you like. The HLA, and/or a Peptide Sequence would be considered minimal.

Sample Input

Input Options

Patient ID		Peptide ID	
Patient HLA			
Peptide Sequence			
Protein Sequence of User's Viral Strain			

Now check the output items you would like to see.

Output Options

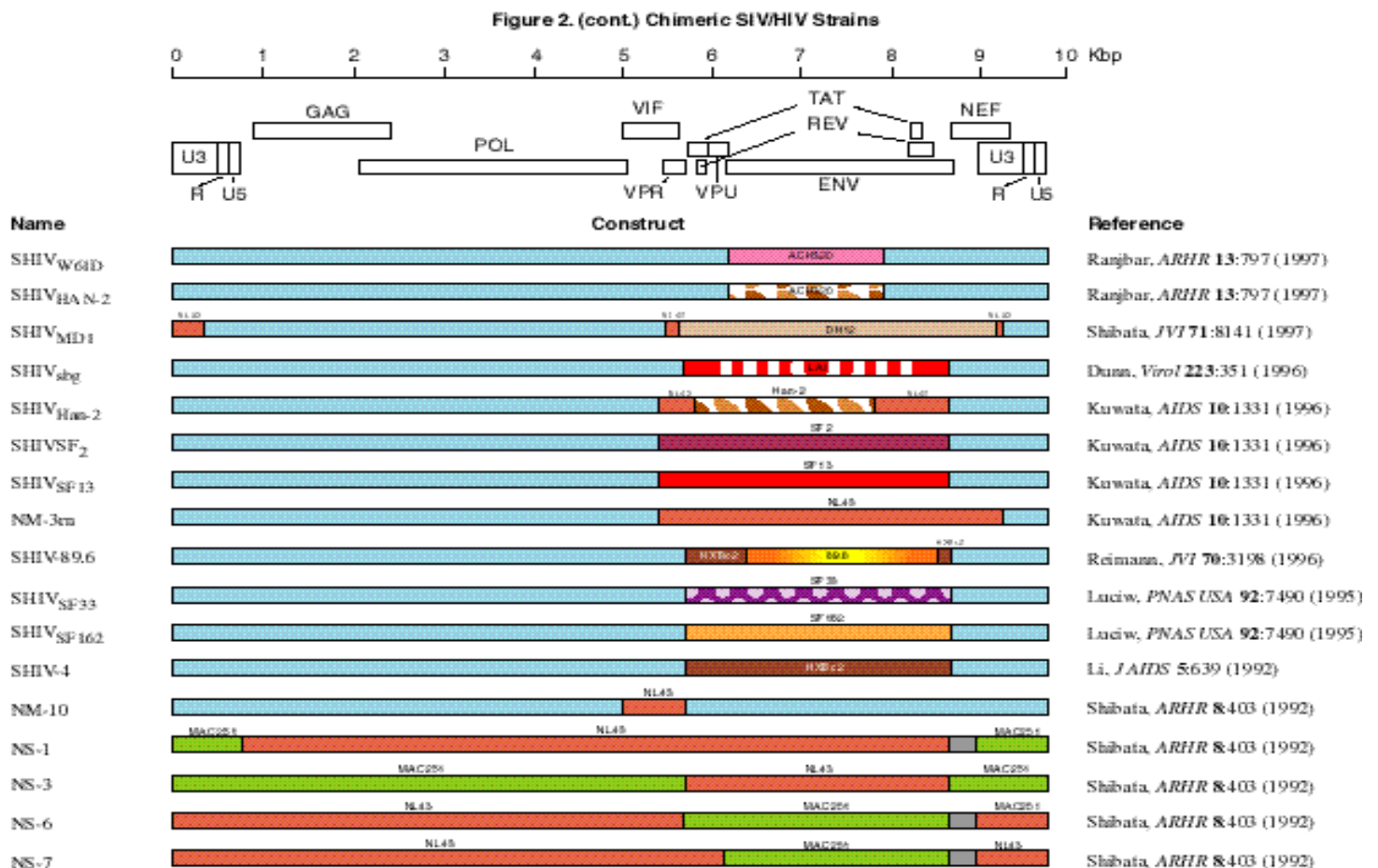
Summary tables	<input checked="" type="checkbox"/>	Find missed epitopes	<input type="checkbox"/>
Align peptide sequence	<input type="checkbox"/>	Be sure to paste in the full protein sequence of your viral strain in the box above.	
Known epitopes in database	<input type="checkbox"/>	Maps	<input type="checkbox"/>

HCV Databases

[Disclaimer/Privacy](#)

SHIV Maps

- SHIVs are Simian (almost always SIV-SMM or SIV-MAC) / Human (usually HIV-1 subtype B) hybrid infectious clones.
- They are frequently used as vaccine reagents and/or challenge virus.
- Many are derived from clones, but often derived via complex in-vivo passage histories.
- Some have been sequenced, but many are only known via sequence of parental lineages.
- Many authors do not adequately describe the SHIVs they are working with, so we have created a review paper to help describe them.
- <http://www.hiv.lanl.gov/content/hiv-db/REVIEWS/VACCINE2001/Vaccine2001.html>



SIV/PIV Alignments

- Any non-human lentivirus is a SIV, not just the SIV-SMM/SIV-MAC group from Sooty mangabeys.
- HIV-1s (M, N and O groups) are related to the SIV-CPZs from the *Pan troglodytes troglodytes* chimps. We describe these alignments as HIV-1/CPZ
- HIV-2s and SIV-MACs are related to SIV-SMMs from Sooty mangabeys. We describe these alignments as HIV-2/SMM
- Dozens of other diverse non-human primates, such as African green monkeys, carry SIVs.
- Alignments of the diverse SIVs, plus HIVs, can help to identify highly conserved codons and other features. We describe these alignments as “other SIV” or HIV-1/HIV-2/SIV.

- http://www.hiv.lanl.gov/content/hiv-db/ALIGN_CURRENT/ALIGN-INDEX.html

The screenshot shows the HIV Sequence Database website. The main content area is titled "2002 HIV and SIV alignments". It features a search bar, a dropdown menu for "Alignment format" set to "FASTA", and a link to "Click here for information on color-coded protein alignments." Below this is a "Get Alignment" button and a "reset" button. A note states: "Note: The protein alignments provided for each gene were constructed using both nucleotide and translated amino acid sequences. Because the translations are based on alignments, they may differ from a straight, non-aligned, translation. For instance, an aligned translation will include frameshift compensation." A table lists various regions and their corresponding alignment options (DNA or Protein).

Region	HIV-1/SIVcpz	HIV-2/SIVsmm	Other SIV
Genome	<input type="radio"/> DNA	<input type="radio"/> DNA	<input type="radio"/> DNA (includes HIV-1 and HIV-2 sequences)
LTR	<input type="radio"/> DNA	<input type="radio"/> DNA	<input type="radio"/> DNA
GAG	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
POL	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
ENV	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
VIF	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
TAT	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
REV	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
VPU/VPX	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
VPR	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein
NEF	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein	<input type="radio"/> DNA <input type="radio"/> Protein

Questions or comments? Contact us at seq-info@t10.lanl.gov

Los Alamos NATIONAL LABORATORY

Operated by the University of California for the US Department of Energy
Copyright © 2001 UC | [Disclaimer/Privacy](#)

NATIONAL INSTITUTES OF HEALTH

HCV Databases

[Disclaimer/Privacy](#)

Vaccine database

- Under construction
- Non-human primate models
- <http://hiv-web.lanl.gov/cgi-bin/vaccine/public/index.cgi>
- Aims: To introduce the new database and to interactively retrieve information related HIV/SIV vaccine studies in nonhuman primates
- Search criteria:
 - ☐ Vaccine
 - ☐ Challenge
 - ☐ Adjuvant
 - ☐ Objective: Immunogenicity, challenge, etc.
- Sources of Databases
 - ☐ LANL
 - ☐ EMMS Corporation (Dr Jon Warren)
- Output:
 - ☐ Reference
 - ☐ Summary
 - ☐ Results
 - ☐ Example of search
 - ☐ Results tabulation

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media Print Download Upload

Address <http://hiv-dev.lanl.gov/cgi-bin/vaccine/public/index.cgi> Go Links

HIV / SIV VACCINE TRIALS DATABASE

(work in progress)

[Search](#) [Clear Form](#) [Help](#) [About This Database](#) [News \(updated 28 Jan, 2004\)](#)

Display Format Order Show results per page

Trial No. Author Title Year

Vaccines

Vaccine Immunogen
HIV-1
HIV-2
SIV
SHIV

Type
Cell/Tissue
DNA
Live attenuated virus

Adjuvant/Stimulant
AS-2 adjuvant
Adju-Phos
Adjuvax

Route
Cervical
Conjunctival
Intra-amniotic

Challenges

Virus
HIV-1
HIV-2
SIV
SHIV

Strain

Route
Cervical
Conjunctival
Intra-amniotic

Species
Cercopithecus aethiops (African Green monkeys)
Macaca (sp)
Macaca fascicularis (cynomolgus macaque)

Objective
Challenge
Chemotherapeutic
Immunogenicity

[Search](#) [Clear Form](#) [Help](#)

Please send questions, comments and suggestions to vaccines-db.lanl.gov

Internet

**Please leave any comments
or suggestions with us:**